DOCUMENT RESUME

ED 199 268 TM 810 164

AUTHOR Green, Donald Ross
TITLE Minimizing Bias in Educational Testing.
PUB DATE 12 Jun 79
NOTE 9p.; Paper presented at the Annual Conference on Large Scale Assessment (9th, Denver, CO, June 12, 1979).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Educational Testing; *Test Bias; *Test Construction; *Test Interpretation
IDENTIFIERS Testing Conditions

ABSTRACT

Sources of test bias are discussed and steps to prevent or reduce bias in tests are listed. Test bias can occur because of the way test materials are written, the conditions of administration, and the interpretations given the results. Steps to prevent or reduce bias arising in the test development process include: (1) using heterogeneous sets of test writers and editors; (2) having test materials reviewed by members of minority groups; (3) limiting the role of language by using performance items; (4) avoiding any items that assume general knowledge not part of the relevant program; and (5) examining item data from tryouts or prior administration separately by ethnic group. Prevention or reduction of bias arising from administrative conditions includes: (1) using examiners familiar to the examinees; (2) making testing situations similar to the learning situation; (3) providing repeated practice tests with feedback; and (4) keeping examiners ethnically heterogeneous. Prevention of bias arising from test scoring, interpretation and use includes: (1) using only objectively scorable measures; (2) training personnel to make legitimate generalizations from test scores; and (3) specifying the intended use of scores. (RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED199268

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Minimizing Bias in Educational Testing

by
Donald Ross Green
CTB/McGraw-Hill

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
D. R. Green
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

The Ninth Annual Conference on
Large Scale Assessment
sponsored by
National Assessment of Educational Progress
June 12, 1979
Denver, Colorado

TM 810 164

Bias in educational assessment procedures is undoubtedly a common phenomenon. It is most widely recognized that this occurs when tests are used as part of the assessment because test data are relatively objective and make it easy to investigate the matter. These very facts probably mean that tests are less biased than other assessment procedures to begin with but whether that is true or not they also make it possible to reduce or minimize the bias.

Bias begins with cultural differences. Thus most, if not all, tests used in schools, particularly those of academic ability and achievement, are biased to some degree for some cultural groups. This is inevitable since schools exist to teach first and foremost the elements of the particular culture - the knowledge, the skills, the attitudes and values - found in that society. It is often possible to reduce the bias in tests but rarely if ever can it be eliminated. The more culturally heterogeneous the society, the greater the problem.

For most groups in American society this problem is probably not large although it must be substantial for a few, notably those who speak little English and those whose membership in our society is very recent. No systematic quantitative indices of these matters exist but it does seem reasonable to infer that in general the problem has been reduced steadily in our society by growing universality of education, increasing mobility, and by the advent of first radio and then television.

Bias arises when the meaning of the scores of a test is systematically different for different groups but is treated as though it was the same (Green, 1975). For example most measures of reading comprehension are based on questions about a passage the examinees read. The task depends on the examinees' familiarity with the content. If prior knowledge of the words and concepts in the passage is related to group memebership, differences in reading skill may be inferred.

inappropriately. As long as this does not occur in an identical fashion for all passages it can be detected in item analyses. This illustrates the point that bias can be reduced or eliminated only if evidence of its existence is obtained. The basic evidence for bias is obtained when separate but parallel validity studies are done for each group. Thus only published tests or those produced to the standards of published tests, i.e., those for which tryout data have been used and validation studies performed, can be improved this way. It is probable that, on the average, informal unpublished tests are the most biased.

Bias may also occur in test use independently of its presence in the test itself. Biased use of any test is possible and occurs when the scores are interpreted or used differentially without reference to any differential validity. Such misuses can be corrected only if evidence about this use is collected - who, when, where, and how all need to be known. Probably most biased uses arise from ignorance and thus training of test users is the most promising if not the only solution.

More specifically bias of any sort can best be dealt with in reference to its possible sources.

Sources of Bias

It can safely be assumed that any bias in tests is unintentional; even strongly bigoted people rarely set out to be unfair. In fact bigotry is not a likely candidate as a source of bias in tests although it could be a part of any of the three listed below.

Basically test bias can occur because of:

(1) The nature of test materials, e.g., the way they are written.

(2) The conditions of administration, e.g., the attitudes of the examiners.

(3) The interpretations given the results, e.g., the hidden assump-

tions made by those using the results.

It seems likely that the first of these categories includes the most probable sources of bias in most tests used in evaluation and curriculum planning, while the second and third tend to become prominent in guidance and selection situations.

1) The nature of test materials can be a source of bias. This can happen if the authors and/or writers and/or editors of the tests differ from some identifiable group of the examinees in their language, styles of thinking, and general knowledge and background and if these elements enter into the test but are irrelevant to its purposes. These elements often enter into written tests more strongly than performance tests since in the latter there is less reliance on language, and verbal skill plays a smaller role in examinee performance. However it should be noted that any item selection or modification procedures based on data from a tryout or a prior administration of the test are likely to accentuate whatever bias of this first sort there is in the test, assuming that the samples have a majority of examinees of the same ethnic group as the test writers. This is because the items that favor this majority group will tend to have data indicating they are better items (Green, 1971).
2) Bias can stem from test administration in several ways. First it can occur when the examiner and examinee come from different ethnic groups. Misunderstanding and/or bias may arise and hurt the performance of examinees if this is an unusual event and/or if the examinees are highly motivated and anxious. Note that while excessive anxiety can hurt performance and can be a biasing event when it impacts one group more than another for reasons irrelevant to the testing,

nevertheless some anxiety in a test situation may improve performance since people trained to take tests can often perform better under pressure. Second, the test situation may be alien to the attitudes and values of some ethnic group whose members may for example, emphasize cooperation rather than competition and demonstrations of individual prowess. Third is the possibility of unfair administration as an unconscious or deliberate expression of personal prejudice on the part of the examiner.

3) Bias can arise in the scoring, interpretation, and use of a test. The most probable way for bias to enter here is from the overgeneralization of the meaning of the results. For example, if one group tends to score lower than another on a particular set of tests some people may impute general inferiority to the low scoring group. Another possible way to introduce bias is to use the test results in ways not intended but which require assumptions about matters not tested. If these assumptions are more reasonable for one group than another the result is bias. Last and presumably least is the deliberate distortion of the results to favor one group over another.

Reducing Bias

Steps to prevent or reduce bias in tests can be inferred from an examination of the possible sources of bias listed above.

1) To reduce the probability of bias arising from the first category of sources a number of steps are possible. These steps all involve steps in the test development process. Reduction of bias from these sources can be accomplished by

   a) Using ethnically heterogeneous sets of test writers and editors. This is most easily done for large-scale development and publishing efforts.

b) Having test materials reviewed by members of minority groups who are sensitive to potential problems of these sorts. This step alone is probably not very useful (Plake, 1979; Sandoval & Whelan, 1979) but can at least avoid overt racism, sexism and other expressions of prejudice.

c) Limiting the role of language, except when the intention is to measure language skills; use performance items whenever possible.

d) Avoiding any items that assume general knowledge not part of the relevant program. This is often very difficult to do.

e) Examining item data from tryouts and/or prior administration separately by ethnic group. Look especially for differences in discrimination indices and in response patterns to distractors in multiple choice items. If possible look at item criterion relationships. Where differences are found, delete or change the item.

Using a latent trait approach, both differential item fits and differences among groups in the parameters of items that do fit can be used to identify and eliminate biased items. Another sound approach similar to the latent trait procedures is to compare item characteristic curves. A variety of ways to do this have been reported. Merz (1978), Merz and Grossen (1979), and Rudner, Getson and Knight (1979) reviewed most of these procedures.

2) The reduction and prevention of bias arising from administrative conditions can be accomplished by

a) Using examiners familiar to the examinees. This would help reduce undue feelings of pressure and improve communication.

b) Making the testing situations very similar to the learning situation which not only would be an anxiety reducing move but which would help those who find test situations alien and/or distasteful.

c) Providing repeated practice tests with feedback on both the substantive matter being tested and on test taking skills. This should reduce the influence of testwiseness on scores.

d) Keeping the examiners ethnically heterogeneous so any interactions would be random events. An alternative would be to match examiners and examinees; this might work as in (a) above but it could also accentuate group differences.

3) Reducing or eliminating possible bias arising from test scoring, interpretation and use can be accomplished by

a) Using only objectively scorable measures.

b) Using machine scorable documents on which examinees record responses where possible.

c) Recording behavior rather than scores when examinee performance must be observed; scoring can then be done independently.

d) Training personnel to distinguish legitimate from inappropriate generalizations and inferences from the test scores.

e) Specifying carefully and publicly the intended uses of the test scores and preventing their use for other unvalidated purposes.

Conclusions

Bias in testing can be reduced only if first one recognizes the possibility of bias in the test, the testing situation and in its use, then second obtains relevant evidence, and finally uses that information to improve the test, the testing situation and the way it is used. Clearly maximum reduction is more

possible with published standardized tests and when the administration, scoring, and score interpretation is left in the hands of qualified professional personnel.

Bibliography

Green, D. R. Racial and ethnic bias in test construction. Monterey, CA: CTB/McGraw-Hill, 1971.

Green, D. R. What does it mean to say a test is biased? Education and Urban Society, 1975, 8 (1), 33-52.

Merz, W. R. Test fairness and test bias: A review of procedures. In M. J. Wargo and D. R. Green (Eds.), Achievement testing of disadvantaged and minority students for educational program evaluation. Monterey, CA: CTB/McGraw-Hill, 1978.

Merz, W. R., & Grossen, N. E. An empirical investigation of six methods for examining test item bias. Final Report (NIE-6-78-0067), Foundation of California State University, Sacramento, 1979.

Plake, B. S. An investigation of a reviewer's ability to identify items which function differentially for selected groups of subjects. Paper presented at the annual meeting of the National Council of Measurement in Education, San Francisco, April 1979.

Rudner, L. M., Getson, P. R., & Knight D. L. The effect of various test and item properties on five approaches to biased item detection. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, April 1979.

Sandoval, J. H., & Whelan, M. P. Accuracy of judgments of WISC-R item difficulty for minority groups. Paper presented at the annual meeting of the American Psychological Association, New York, September 1979.